Problema II

Dist	A	B	C	D	E
A	0	1	2	9	10
B		0	3	7	5
C			0	4	6
D				0	8
E					0

Dada la siguiente matriz de distancias entre cinco patrones, se pide:

1. Aplicar un clustering jerárquico con el método de enlace simple:

Calculemos primero las distancias en cada iteración:

Iteración 1*

$A \equiv C_1$; $B \equiv C_2$; $C \equiv C_3$; $D \equiv C_4$; $E \equiv C_5$

Dado que las distancias para la primera iteración estan representadas en la tabla, no se volverán a calcular:

Iteración 2

$C_1 = (A, B)$; $C_2 = C$; $C_3 = D$; $C_4 = E$

$d(C_1, C_2) = \min\left(d(A, C), d(B, C)\right) = \min(2, 3) = 2$

$d(C_2, C_3) = \min(d(C, D)) = 4$ | $d(C_3, C_4) = \min(d(D, E)) = 8$

$d(C_2, C_4) = \min(d(C, E)) = 6$

Tomando los cálculos en cuenta se tendrá que

$C_1 = (A, B, C)$; $C_2 = D$; $C_3 = E$

Iteración 3

$C_1 = (A, B, C)$; $C_2 = D$; $C_3 = E$

$d(C_1, C_2) = \min(d(A, D), d(B, D), d(C, D)) = \min(9, 7, 4)$

$d(C_1, C_2) = 4$

$d(C_1, C_3) = \min(d(A, E), d(B, E), d(C, E)) = \min(10, 5, 6)$

$d(C_1, C_3) = 5$

$d(C_2, C_3) = \min(d(D, E)) = 8$

Por lo que la tercera iteración tendremos los siguientes clústers:

$C_1 = (A, C, B, D)$; $C_2 = E$

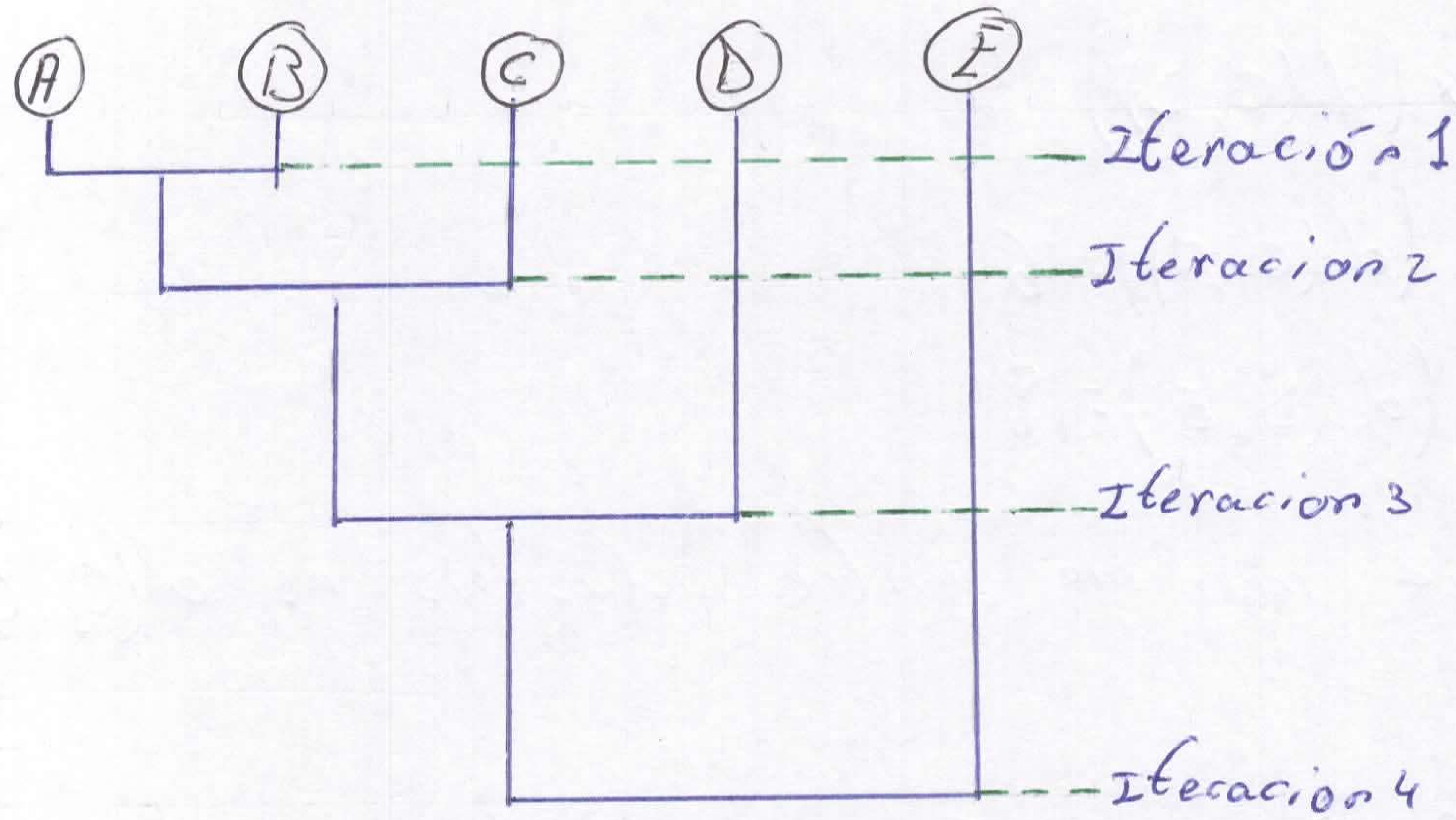
Iteración 4

$C_1 = (A, B, C, D)$; $C_2 = E$

$d(C_1, C_2) = \min(d(A, E), d(B, E), d(C, E), d(D, E)) = (10, 5, 6, 8) = 5$

Con esto se han completado todas las iteraciones.

Realicemos ahora las representaciones gráficas:



2.- Aplicar un clústering jerárquico con el método de enlace complejo.

Empecemos por calcular las distancias entre clústers para cada iteración

Iteración 1

$C_1 = A$, $C_2 = B$, $C_3 = C$, $C_4 = D$, $C_5 = E$

Al igual que en el ejercicio anterior, el calculo de todas las distancias representa reescribir la tabla de distancias por lo que obbiaremos este paso

Iteración 2

$C_1 = (A, B)$, $C_2 = C$, $C_3 = D$, $C_4 = E$

$d(C_1, C_2) = max(d(A,C), d(B,C)) = max(3,2) = 3$

$d(C_1, C_3) = max(d(A,D), d(B,D)) = max(9,7) = 9$

$d(C_1, C_3) = max(d(A,E), d(B,E)) = max(10,5) = 10$

$d(C_2, C_3) = max(d(C, D)) = 4$

$d(C_2, C_4) = max(d(C, E)) = 6$

$d(C_3, C_4) = max(d(D, E) = 8$

La distancia minima entre clusters es $d(C_1, C_2)$

Iteración 3

$C_1 = (A, B, C)$, $C_2 = D$, $C_3 = E$

$d(C_1, C_2) = max(d(A, D), d(B, D), d(C, D) = max(9, 7, 4)$

$d(C_1, C_2) = 9$

$d(C_1, C_3) = max(d(A, E), d(B, E), d(C, E)) = max(10, 5, 6)$

$d(C_1, C_3) = 10$

$d(C_2, C_3) = max(d(D, E)) = 8$

La distancia minima es la distancia entre los clúster C_2 y C_3 que equivale a $\underline{8}$

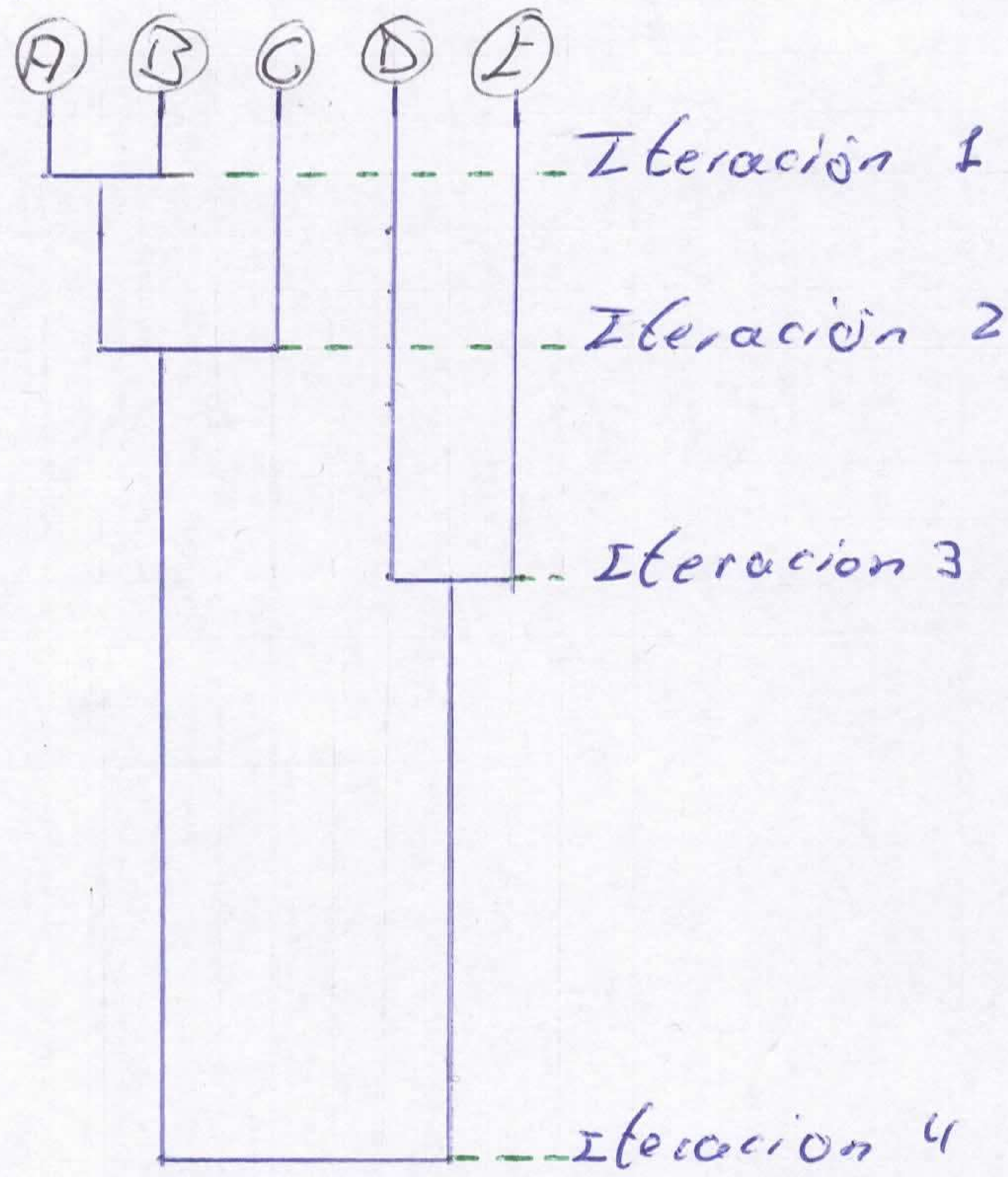
Iteracion 4

$C_1 = (A, B, C)$, $C_2 = (D, E)$

$d(C_1, C_2) = max(d(A, D), d(B, D), d(C, D), d(A, E), d(B, E), d(C, E))$

$d_1(C_1, C_2) = max(9, 7, 4, 10, 5, 6) = 10$

Representemos gráficamente los resultados:



3.- ¿Se podría aplicar el método de enlace medio con estos datos? Justifica la respuesta.

Sí, el método de enlace medio tiene un problema y es el alto coste computacional que conlleva ya que hay que calcular la distancia media entre cada par de puntos de todos los puntos pero al ser en este caso menos puntos no representa un problema ni diferencia en el coste computacional. Pese a esto el método de enlace medio se puede aplicar en todas las situaciones en las que se puedan emplear los métodos ya usados.